DECS 431 – Sample Waiver Exam

Estimating Rents in the Residential Real Estate Market

This case studies how the rent for an apartment is related to the characteristics of the apartment. For this purpose we look at a sample of rental rates for one-, two-, and three-bedroom apartments in the Los Angeles area. All relevant information is contained in the file LA.xls where the variables are:

RENT   =   monthly rental in dollars

COMMON = total number of common rooms (Rooms that are kitchens, living rooms, or dining areas are classified as common rooms. Note that this number may be fractional, as rooms such as living/dining area combinations may be counted as more than one but less than two rooms.)

SQKLD   =   total square footage of the common rooms

BED   =   number of bedrooms

SQBED   =   total square footage of bedrooms

BATH   =   number of bathrooms

(Note that this number may be fractional, since, for example, a bathroom with only a sink and toilet but no shower or bath counts as only 0.5 a bathroom.)

SQBATH =   total square footage of bathrooms

PKG   =   number of parking spaces included with the rent

BEACH   =   number of miles from the beach

UCLA =   number of miles from the UCLA (University of California, Los Angeles) campus

We will use these data to estimate what is called an "hedonic" rent model. Similar models have been calibrated in attempts to explain the selling prices of properties in the residential and commercial real estate markets. The idea is to use regression analysis to decompose the rental rate (or selling price) into components due to different features of the property. Categories of variables often used include the features of the dwelling, the lot, the neighborhood, or the

Kellogg
Department of Managerial Economics and Decision Sciences
Kellogg School of Management n Northwestern University

DECS-431

municipality. One use for such a model is to predict the rental rate or selling price for a given property. This is an "assessment" strategy such as that used by realtors or mortgage lenders. Table 1 below shows a regression with rent as the dependent variable and 9 independent variables.

Table 1

Regression: rent

	Constant	common	sqkld	bed	sqbed	bath
Coefficient	435.9384	-26.8466	0.952491	44.269	1.143276	-23.47182
std error of coef	81.84318	25.21427	0.152702	17.44241	0.170228	50.92596
t-ratio	5.3265	-1.0647	6.2376	2.5380	6.7161	-0.4609
p-value	0.0001%	29.0654%	0.0000%	1.3377%	0.0000%	64.6299%
beta-weight		-0.0257	0.3353	0.1554	0.3908	-0.0403

	sqbath	pkg	beach	ucla
Coefficient	2.049392	23.27649	-17.6021	-4.03163
std error of coef	1.391717	11.04677	2.811051	2.377619
t-ratio	1.4726	2.1071	-6.2617	-1.6957
p-value	14.5351%	3.8695%	0.0000%	9.4394%
beta-weight	0.1342	0.0725	-0.1453	-0.0420

Standard error of regression	49.63199
R-squared	96.33%
Adjusted R-squared	95.85%
number of observations	80
residual degrees of freedom	70
t-statistic for computing 95%-confidence intervals	1.9944

For Questions 1-5, please use the regression in Table 1 as your model. Note that you may need to do further calculations/analysis of this regression with Excel/Kstat.

QUESTION 1

Suppose a landlord owns an apartment with three common rooms, one bedroom, one bathroom, and one parking space. It has 300 square feet total in the common rooms, 45 square feet in the bedroom, and 40 square feet in the bathroom. It is 4 miles from the beach, and 5 miles from the UCLA campus. What rent do you expect this landlord to charge? Also, provide an interval that you are 95% confident contains the rent for this apartment. What is the estimated probability that rent on an apartment with these characteristics would be more than $800 per month?

QUESTION 2

A landlady owns an apartment located in the same apartment complex as the apartment we examined in question 1. Her apartment has two bedrooms that are 45 square feet each, and is otherwise identical to the apartment considered in question 1. How much more do you expect her to charge in rent for her apartment compared to the rent for the apartment in question 1?

QUESTION 3

Which (if any) of the variables in the Regression in Table 1 seem like they may significantly affect rent? (Use a 10% level of significance as your standard.)

QUESTION 4

Common wisdom among the realtors in the LA area says that every additional mile away from the beach reduces the rent of an apartment by more than $25 per month. Can you reject this claim using a 5% level of significance?

QUESTION 5

The landlady discussed in question 2 learns of your regression analysis (in Table 1). Being quite bright, she notices that increasing the number of bedrooms seems to result in higher rents. As a result, she remodels her apartment described in question 2 by taking the two existing bedrooms and subdividing each into three bedrooms! (for a new total of six bedrooms).

(a) What is the predicted increase in rent that will result from her remodeling?
(b) Provide an interval that you are 90% confident contains the increase in rent due to her remodeling.

Kellogg
Department of Managerial Economics and Decision Sciences
Kellogg School of Management n Northwestern University

DECS-431

(c) She actually does go ahead with the remodeling. However, the rent she is able to get for the remodeled apartment turns out to be substantially less than the model in Table 1 predicts (and even less than the lower end of the 95% prediction interval) and she is very disappointed. Looking at the data and your regression model (i.e., don't just tell me that the bedrooms were too small!), why might we expect the model's prediction could be wrong in this case?

QUESTION 6

A realtor (trying to save money) did not purchase the full set of data. He only bought the data listing the rent, the number of common rooms, the number of bedrooms, and the number of bathrooms. Using only this data, what is the best regression model for this realtor to use?

(a) Write down the estimated regression equation.

(b) Describe in a concise manner how you arrived at this regression equation.

(c) Use your new regression to estimate the average rent of all apartments that have 3 common rooms, 3 bedrooms, and 2 bathrooms. Please describe explicitly how you did your calculation.

Kellogg
Department of Managerial Economics and Decision Sciences
Kellogg School of Management n Northwestern University

DECS-431

Solutions

Question 1: We plug in the values in the prediction worksheet:
Prediction, using most-recent regression

	constant	common	sqkld	bed	sqbed	bath	sqbath	pkg	beach	ucla
Coefficients	435.9384	-26.8466	0.9525	44.27	1.143	-23.47	2.049	23.28	-17.6	-4.032
values for prediction			3	300	1	45	1	40	1	4	5

predicted value of rent	728.0762
standard error of prediction	52.20081
standard error of regression	49.63199
standard error of estimated mean	16.17377

confidence level	95.00%
t-statistic	1.9944
residual degr. Freedom	70

confidence limits	**lower**	623.9651
for prediction	**upper**	832.1874

confidence limits	**lower**	695.8187
for estimated mean	**upper**	760.3338

a) $728.08.
b) The 95% prediction interval is ($623.97, $823.19).
c) We should use the standard deviation of prediction, which is 52.2. Normalizing $800 we get
$$\text{t-value} = (800 - 728.08) / 52.2 = 1.378,$$
and the corresponding p-value is TDIST(1.378,70,1) = 8.6%.

Question 2: The effect of one additional bedroom that is 45 square feet is
$$1 * 44.269 + 45 * 1.1432 = \$95.716.$$

Question 3:
The p-values of **sqkld**, **bed**, **sqbed**, **pkg**, **beach** and **ucla** are below 10%, hence these parameters are significant at a 10% level of significance.
For the remaining three parameters, **common**, **bath** and **sqbath**, we should check for a possible multicollinearity problem.
The variance inflation factors are:

variance inflation								
1.1097826	5.5037194	7.1386758	6.4510622	14.587352	15.819966	2.2534069	1.0264721	1.1667706
common	**Sqkld**	**bed**	**sqbed**	**bath**	**sqbath**	**pkg**	**beach**	**ucla**

The VIF of **common** is low, so there is no multicollinearity problem with that variable - it is not significance at a 10% level of significance.

However, there is a multicollinearity problem with **bath** and **sqbath**.
We check for joint significance, and get

Analysis of variance

	base model		extended model		difference	
	sum of squares	df	sum of squares	df	sum of squares	df
Regression	4506260.115	7	4520481.311	9	14221.19609	2
Residual	186654.5729	72	172433.3768	70	172433.3768	70
Total	4692914.688	79	4692914.688	79	186654.5729	72
F-ratio	248.3202		203.9008		2.8866	
degrees of freedom	(7, 72)		(9, 70)		(2, 70)	
p-value	0.00000%		0.00000%		6.24302%	

The p-value is 6.24%, so at a 10% level of significance, at least one of them is significant.
Since we cannot know which one is the significant one (or maybe both are), we must say that
both might be statistically significance at a 10% level of significance.

Question 4: The claim is equivalent to: the coefficient of beach is less than -25.
We therefore set up the following hypothesis test:
H_0 : coefficient of beach $\leq$ -25.
H_A : coefficient of beach > -25.
Our estimator is -17.6, and its standard deviation is 2.811.
Normalizing the estimator we get
$$\text{t-value} = (-17.6 - (-25)) / 2.811 = 2.63,$$
and the corresponding p-value is TDIST(2.63, 70, 1) = 0.52%.
This is below 5%, so we can accept the alternative and reject the null hypothesis.
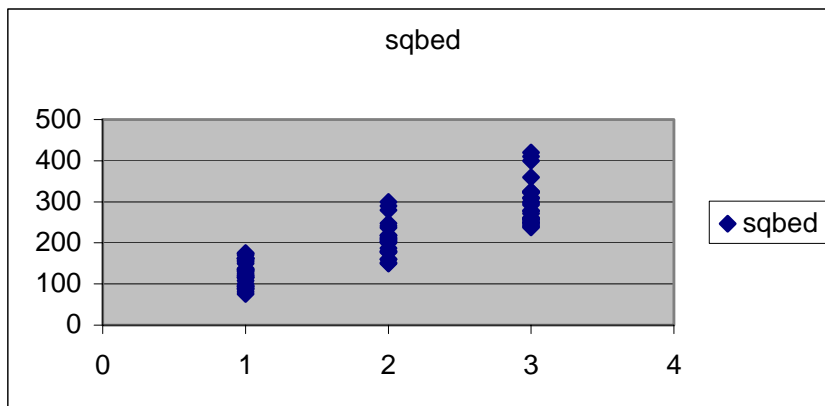
Question 5:
a) Since she adds 4 bedrooms to her apartment, without changing the total area of the bedrooms,
    the predicted increase is 4 * 44.269 = $177.08.
b) A 90% confidence interval for the coefficient of **bed** is
$$44.269 \pm \text{TINV}(0.1,70) * 17.44 = (15.19, 73.34).$$
Hence, the desired interval is (4 * 15.19, 4 * 73.34) = (60.78, 293.37).

c) The scatter plot of **bed** and **sqbed** looks as follows:

In particular, we have no data on apartments with more than 3 bedrooms. Moreover, we have no data on apartments where the average size of the bedrooms is 15 square feet.
Thus, we do not know whether the relation remains linear for such apartments, and, in particular, the regression cannot be used to provide useful predictions in such cases.

Question 6:
A linear model looks like this:

Regression: rent

	constant	common	bed	Bath
Coefficient	603.321872	-61.203846	192.257131	193.923743
p-value	0.0014%	17.0944%	0.0000%	0.0000%

The Breusch-Pagan index is 3.6%, which suggests that there is a heteroskedasticity problem.

Next, we try a semi-log model, by adding a **ln(rent)** variable.
The model is
Regression: ln(rent)

	constant	common	bed	Bath
Coefficient	6.51505805	-0.0584273	0.18714243	0.17171577
p-value	0.0000%	17.0539%	0.0000%	0.0001%

The Breusch-Pagan index is 52%, which suggests that there is no heteroskedasticity problem. The residual plot looks as follows:

Residual Plot

predicted values of ln(rent)
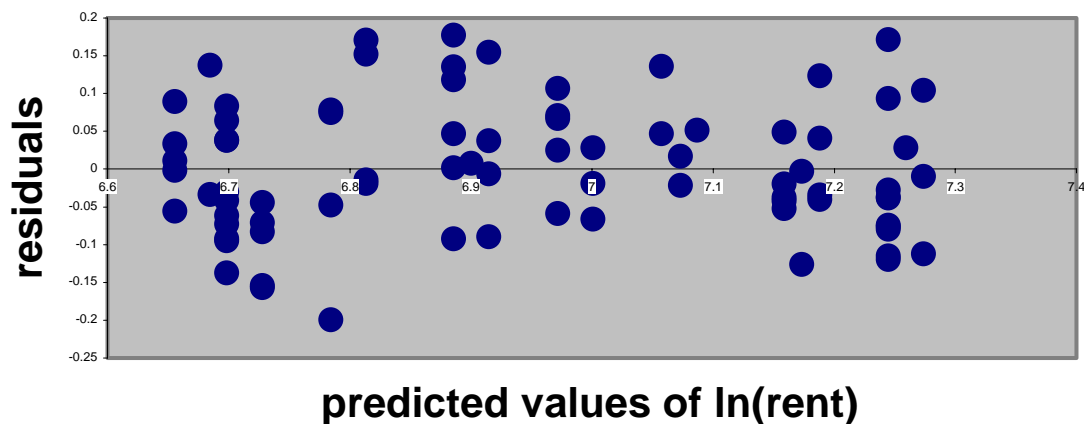
We see no evident patterns, so there is no evidence for non-linear relation.

The p-value of **common** is 17% and the variance inflation factor is low, so we drop this variable.

The new model is:
Regression: ln(rent)

Kellogg   DECS-431
Department of Managerial Economics and Decision Sciences
Kellogg School of Management n Northwestern University

	constant	bed	bath
Coefficient	6.35028602	0.1862321	0.1728118
p-value	0.0000%	0.0000%	0.0000%

The Breusch-Pagan index is 59%, the residual plot looks random, and so we adopt this model.

We also check the log-log model.
Here the regression equation is:
Regression:
ln(rent)

	constant	ln(common)	ln(bed)	ln(bath)
Coefficient	6.83645466	-0.1333195	0.33943814	0.22909178
p-value	0.0000%	24.1300%	0.0000%	0.0000%

The Breusch-Pagan index is 36%, and the residual plot does not show any evident patterns. Since the p-value of **ln(common)** is high, we drop it and get a more compact model:
Regression:
ln(rent)

	constant	ln(bed)	ln(bath)
Coefficient	6.69882963	0.33788375	0.23106524
p-value	0.0000%	0.0000%	0.0000%

with Breusch-Pagan index 38% and no evident patterns in the residual plot.
So this is a valid model as well.

c) We use the semi-log model
Regression: ln(rent)

	constant	bed	bath
Coefficient	6.35028602	0.1862321	0.1728118

The prediction worksheet gives us a predicted value of
$$\mathbf{ln(rent)} = 6.35 + 3 * 0.186 + 0.172 * 2 = 7.25.$$
To get a prediction of the average **rent**, we should exponentiate 7.25, and multiply by the correction factor:
$$\mathbf{rent} = \exp(7.25) * \exp(0.088^2 / 2) = 1414 * 1.00388 = \$1420.$$